

⑩ 日本国特許庁 (J P)

⑪ 特許出願公開

⑫ 公開特許公報 (A)

昭60-75900

⑬ Int. Cl.<sup>4</sup>  
G 10 L 9/08

識別記号  
G L A

庁内整理番号  
7350-5D

⑭ 公開 昭和60年(1985)4月30日

審査請求 未請求 発明の数 1 (全6頁)

⑮ 発明の名称 単語音声認識装置

⑯ 特 願 昭58-183841

⑰ 出 願 昭58(1983)9月30日

⑱ 発 明 者 下 谷 光 生 尼崎市塚口本町8丁目1番1号 三菱電機株式会社応用機器研究所内

⑲ 発 明 者 日 比 野 昌 弘 尼崎市塚口本町8丁目1番1号 三菱電機株式会社応用機器研究所内

⑳ 発 明 者 嶋 憲 司 尼崎市塚口本町8丁目1番1号 三菱電機株式会社応用機器研究所内

㉑ 出 願 人 三菱電機株式会社 東京都千代田区丸の内2丁目2番3号

㉒ 代 理 人 弁理士 大 岩 増 雄 外2名

明 報 書

1. 発 明 の 名 称

単語音声認識装置

2. 特 許 請 求 の 範 囲

(1) 入力された音声信号の特徴パラメータを抽出し、該抽出した特徴パラメータと予め登録された複数の単語音声の特徴パラメータとの類似度を計算して単語音声の認識処理を行なう単語音声認識装置において、

前記認識処理を行なう音声信号区間を規定するために、入力された音声信号の始終端を検出する始終端検出手段を備え、

前記始終端検出手段は、

入力された音声信号の自己相関関数の予め定められた範囲内での最大値 CORMAX を計算する CORMAX 計算手段と、

前記 CORMAX に基づいて、音声信号の始終端の決定を行なう始終端決定手段とを含むことを特徴とする、単語音声認識装置。

(2) 前記始終端決定手段は、前記 CORM

AX が予め設定された CORMAX のしきい値を越えたことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第1項記載の単語音声認識装置。

(3) 前記始終端決定手段は、前記 CORMAX が予め設定された CORMAX のしきい値より小さい入力波形が予め設定された区間だけ続いたことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第1項または第2項記載の単語音声認識装置。

(4) 前記始終端検出手段は、入力された音声信号の波形の大きさに対応する量を計算するレベル計算手段をさらに含み、

前記始終端決定手段は、前記 CORMAX と前記レベル計算手段の出力とに基づいて、音声信号の始終端の決定を行なう、特許請求の範囲第1項記載の単語音声認識装置。

(5) 前記レベル計算手段は、入力された音声信号のパワーを計算する手段を含む、特許請求の範囲第1項記載の単語音声認識装置。

(6) 前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値を超え、かつ前記CORMAXが予め設定されたCORMAXのしきい値を超えたことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第5項記載の単語音声認識装置。

(7) 前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値を超え、かつ前記CORMAXに予め設定された定数を掛けた値がパワーより大きいことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第5項記載の単語音声認識装置。

(8) 前記始終端決定手段は、前記CORMAXが予め設定されたCORMAXのしきい値より小さいことおよび前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値より小さいことの少なくともいずれか一方の条件を満たす入力波形が予め設定された区間だけ続いた

ことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第5項記載の単語音声認識装置。

(9) 前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値より小さいことおよび前記CORMAXに予め設定された数を掛けた値がパワーより小さいことの少なくともいずれか一方の条件を満たす入力波形が予め設定された区間だけ続いたことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第5項記載の単語音声認識装置。

### 3. 発明の詳細な説明

#### [発明の技術分野]

この発明は、単語音声認識装置に関し、特にたとえば単語音声区間の始終端検出の改良に関する。

#### [従来技術]

第1図は従来の単語音声認識装置の一例を示す概略ブロック図である。図において、マイクロホン11から入力された音声信号は、マイクロホン

アンプ12で増幅された後、AGC回路13に与えられる。このAGC回路13は、入力信号の大きさが変動しても、一定出力が得られるように、その内部に備えられた増幅器の利得を自動的に制御する回路である。AGC回路13の出力は、A/D変換回路14に与えられ、デジタル信号に変換される。A/D変換回路14の出力は、波形メモリ15に与えられる。この波形メモリ15は、1フレームの入力波形データを一時記憶するメモリである。波形メモリ15の出力は、パワー計算回路21に与えられるとともに、特徴抽出部3に与えられる。パワー計算回路21は、波形のパワー(電力)を計算する回路である。パワー計算回路21の出力は認識処理部6に与えられるとともに、始終端検出回路22に与えられる。始終端検出回路22は、音声信号の始終端を検出する回路であり、その出力は認識処理部6に与えられる。一方、特徴抽出部3はデジタルフィルタなどを含んで構成され、入力音声波形の特徴パラメータを抽出する回路である。特徴抽出部3の出力は認

識処理部6に与えられる。この認識処理部6には、入力パターンメモリ4および登録パターンメモリ5が接続される。入力パターンメモリ4は、単語音声の認識モードにおいて、特徴抽出部3で分析抽出された認識すべき音声の特徴パラメータを一時記憶するメモリである。登録パターンメモリ5は、登録モードにおいて、分析抽出された登録語の特徴パラメータあるいは標準音声の特徴パラメータを予め記憶するメモリである。認識処理部6は、たとえばマイクロプロセッサやマイクロコンピュータなどを含んで構成され、入力パターンメモリ4と登録パターンメモリ5内の特徴パラメータを用いて認識処理を行なう回路である。このような単語音声認識装置においては、音声区間をフレームと呼ばれる一定時間間隔に分割してフレームごとに音声の特徴抽出が行なわれる。

次に、第1図の回路の動作を説明する。マイクロホン11から入力された音声信号はマイクロホンアンプ12、AGC回路13、A/D変換回路14をって一旦波形メモリ15に記憶される。

特徴抽出部3は波形メモリ15から1フレーム分の波形データを受取り特徴パラメータの抽出を行なう。得られた特徴パラメータは、登録モードにおいては登録パターンメモリ5に記憶される。一方、認識モードにおいては、得られた特徴パラメータは一旦入力パターンメモリ4に記憶され、その後認識処理部6でパターンマッチング等の手法により認識処理が行なわれる。

一方、始終端検出回路22は、パワー計算回路21が計算する音声信号のパワーにもとづいて、音声信号区間の始終端を検出する。認識処理部6は、この始終端検出回路22で規定される区間の音声信号を認識すべき音声信号として認識処理を行なう。

第2図は音声信号のパワー波形を示す図である。この第2図を参照して、第1図に示す始終端検出回路22の動作を説明する。始終端検出回路22は、音声信号のパワーが予め設定されたしきい値 $P_s$ を超えると音声信号の始端を検出し、パワーが予め設定されたしきい値 $P_e$ 以下であるフレー

ムが予め設定された区間 $K_{th}$ だけ続くと音声信号の終端を検出する。この例では、 $k_1$ と $k_2$ がそれぞれ音声信号の始端フレームと終端フレームである。前述のように、認識処理部6は始終端検出回路22で規定される音声信号区間すなわち $k_1 \sim k_2$ の区間を認識すべき単語音声として認識処理する。したがって、このような認識装置においては、音声信号の始終端検出の性能が認識結果に大きな影響を与える。

第3図は音声に騒音が加わった場合の音声信号のパワー波形を示す図である。この波形の音声信号区間は第2図に示すように、 $k_1 \sim k_2$ であるにもかかわらず、従来の装置の始終端検出方法では $k_3 \sim k_4$ が音声信号区間であると検出する。このように、従来の音声認識装置は、騒音が強い環境下においては、単語音声の始終端検出が正確に行なわれず、認識性能が下がるという欠点があった。

#### [発明の概要]

この発明は、上述のような従来の装置の欠点を

除去するためになされたもので、単語音声の始終端を予め設定された範囲内の自己相関係数の最大値(以下CORMAXと称す)を用いて検出することにより騒音が大きい環境下においても始終端検出を正確に行ない得て、認識性能の優れた音声認識装置を提供することを目的とする。

#### [発明の実施例]

第4図はこの発明の一実施例を示す概略ブロック図である。図において、この第4図の実施例は、以下の点を除いて第1図の回路と同様であり、相当する部分には同様の参照番号を付しその説明を省略する。この第4図の実施例では、CORMAX計算回路23が設けらる。このCORMAX計算回路23は、たとえば乗算器や加算器からなる自己相関器を含んで構成され、波形メモリ15から入力される音声信号波形のCORMAXを計算する。計算されたCORMAXは、始終端検出回路220の一方入力に与えられるとともに、認識処理部6に与えられる。始終端検出回路220の他方入力には、パワー計算回路21の出力が与え

られる。すなわち、この実施例の特徴は、パワー計算回路21で計算されたパワーとCORMAX計算回路23で計算されたCORMAXとに基づいて、音声信号の始終端を検出することである。

次に、CORMAXについて説明する。1フレーム分の波形データを $x(i)$ 、( $i=1, 2, \dots, I$ )とするとパワー $P$ は次式(1)で表わされる。

$$P = \sum_{i=1}^I x(i) \cdot x(i) \quad \dots (1)$$

$\tau$ 次の自己相関係数 $COR(\tau)$ は次式(2)で表わされる。

$$COR(\tau) = \frac{1}{I-\tau} \sum_{i=1}^{I-\tau} x(i) \cdot x(i+\tau) \quad \dots (2)$$

CORMAXを求めるために設定した自己相関係数の区間を次数 $\tau_s \sim \tau_e$  ( $\tau_s \sim \tau_e$ )とするとCORMAXは次式(3)で表わされる。

$$CORMAX = MAX[COR(\tau)] \quad \dots (3)$$

パワーの大きさが同じ波形であっても母音などのピッチ性の強い波形はCORMAXは大きく、白色雑音に近い環境騒音などの波形はCORMAX

は小さい。この発明は、このことを利用して音声信号の始終端検出を行なうものである。

第5図は音声に騒音を加えた場合のCORMAX波形を示す図であるが、図示のようにCORMAX波形では騒音の影響が緩和されている。したがって、第1図の回路と同様にしきい値弁別を行なって始終端を検出した場合、音声信号区間は $k_5 \sim k_6$ となり、従来の装置に比べて正確に始終端の検出を行なうことができる。この発明では、CORMAXのみに基づいて単語音声の始終端を検出するようにしてもよい。しかしながら、始終端の検出の要素として、CORMAXだけでなく音声信号のパワーやレベルなど波形の大きさに対応する量を組合わせて用いると、さらに正確な始終端の検出が行ない得る。そこで、第4図の実施例では、始終端の検出の要素として音声信号のパワーと、CORMAXとを用いている。

すなわち、第4図の実施例では、音声信号のパワーが予め設定されたしきい値 $P_s$ 以上でしかもCORMAXに予め設定された定数 $C_s$ を掛けた

値がパワー以上であるフレームを音声信号の始端フレームとし、パワーが予め設定されたしきい値 $P_e$ 以下であるかCORMAXに予め設定した定数 $C_e$ を掛けた値がパワー以下であるかの少なくとも一方を満足するフレームを無音フレームとし、この無音フレームが予め設定されたフレーム数 $K_{th}$ だけ続くと、音声信号の終端を検出し、最初の無音フレームを音声信号の終端フレームとしている。

次に、第4図の実施例のさらに詳細な動作を説明する。パワー計算回路21およびCORMAX計算回路23は、波形メモリ15から1フレーム分の波形データを受取り、それぞれ、第(1)式および第(2)式の計算を行ない、パワーとCORMAXを始終端検出回路220に送る。始終端検出回路220は、パワーとCORMAXを用いて音声信号の始終端判定を上述の方法によって行ない、その結果を認識処理部6に与える。認識処理部6では、始終端検出回路220によって検出された始端から終端までの間の音声信号を入力パターンメモリ4あるいは登録パターンメモリ5に

格納し、認識処理を行なう。なお、その他の動作は、第2図に示す従来装置と同様である。

なお、他の実施例として、始終端検出回路220における始終端の検出は、パワーが予め設定されたしきい値 $P_s$ 以上でかつCORMAXが予め設定されたしきい値 $C_{Ms}$ 以上であるフレームを始端フレームとして検出し、パワーが予め設定されたしきい値 $P_e$ 以下であるかCORMAXが予め設定されたしきい値 $C_{Me}$ 以下であるか少なくとも一方を満足するフレームを無音フレームとし、この無音フレームが予め設定されたフレーム数 $K_{th}$ だけ続くと終端を検出し最初の無音フレームを音声信号の終端フレームとするようにしてもよい。この場合も第4図の実施例と同様の効果を奏することはもちろんである。

さらに他の実施例として、始端を検出する場合、パワーの値が1フレーム前のパワーの値よりも大きいという条件を加えてもよく、この場合は始端検出能力を向上することができる。

また、上述の実施例では、始終端を検出するた

めの1要素として音声信号のパワーを用いるようにしているが、このパワーに代えてその他音声信号の波形の大きさを表わす量(波形のレベルなど)を計算して始終端検出のための要素として用いるようにしてもよい。

また、上述の実施例では、パワーが予め設定されたしきい値 $P_s$ 以上でかつCORMAXに予め設定された定数 $C_s$ を掛けた値がパワー以上であるフレームを音声信号の始端フレームとしたが、始端フレームはこの近傍のフレームにしても差し支えない。

また、上述の実施例では、音声信号の終端フレームを最初の無音フレームとしたが、終端フレームはこの近傍のフレームでも差し支えない。

さらに、上述の実施例では、説明の都合上単語音声認識装置を特定話者登録型としたが、予め標準音声の特徴を登録パターンメモリに登録している不特定話者用の単語音声認識装置であってももちろんよい。

[発明の効果]

以上のように、この発明によれば、CORMAXに基づいて単音音声の始終端を検出するようにしたので、騒音が大きい環境下でも音声信号の始終端の検出を正確に行なうことができ、音声認識装置の認識性能を高めることができる。

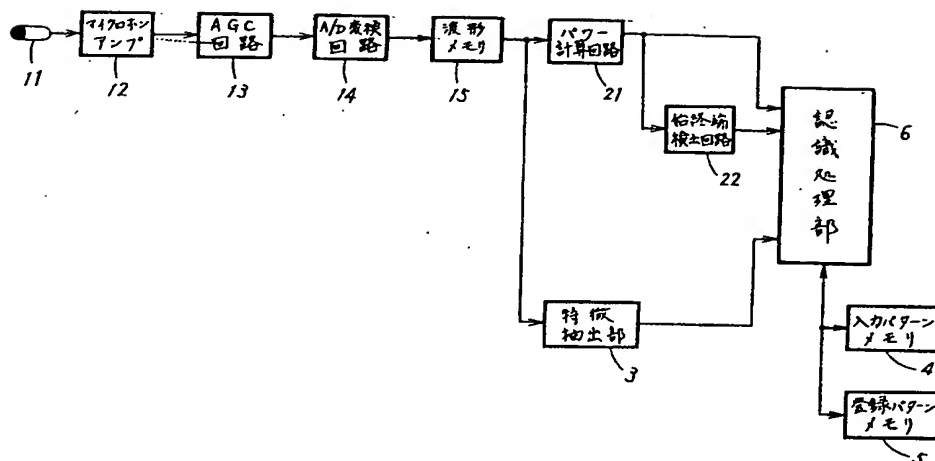
#### 4. 図面の簡単な説明

第1図は従来の単音音声認識装置の一例を示す概略ブロック図である。第2図は音声信号のパワー波形を示す図である。第3図は音声に騒音が加わった場合のパワー波形を示す図である。第4図はこの発明の一実施例を示す概略ブロック図である。第5図は音声に騒音を加えた場合のCORMAX波形を示す図である。

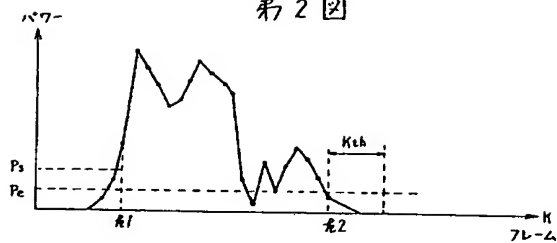
図において、3は特徴抽出部、4は入力パターンメモリ、5は登録パターンメモリ、6は認識処理部、11はマイクロホン、21はパワー計算回路、23はCORMAX計算回路、220は始終端検出回路を示す。

代理人 大 岩 増 雄

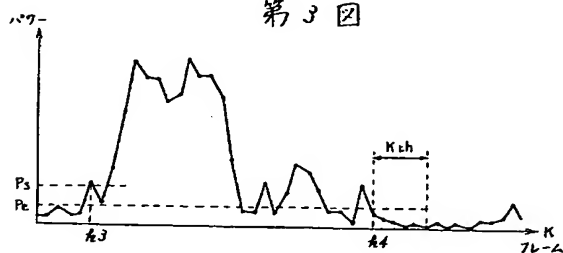
第1図



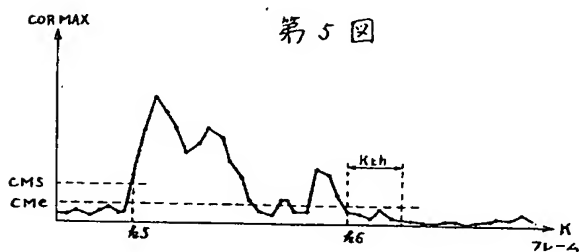
第2図



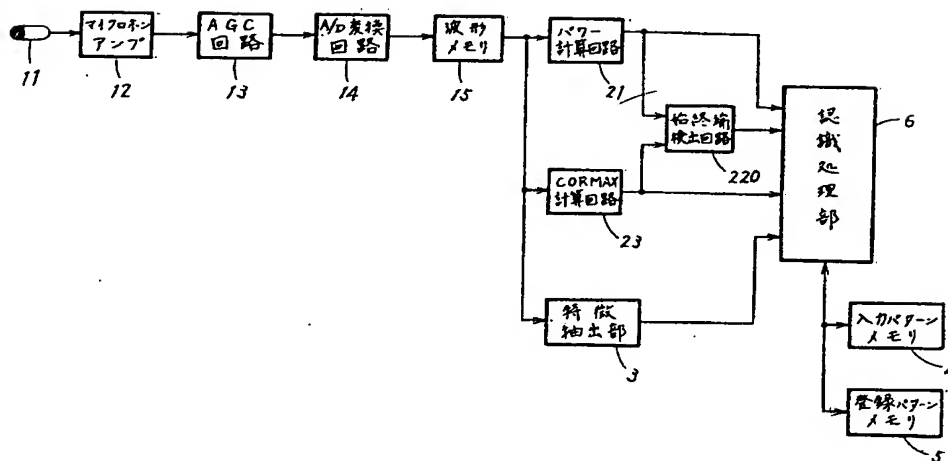
第3図



第5図



第4図



**PARTIAL TRANSLATION OF JAPANESE UNEXAMINED PATENT  
PUBLICATION (KOKAI) NO. S60-075900**

Title of the Invention: A Speech Recognition Apparatus

Publication Date: April 30, 1985

Patent Application No.: S58-183841

Filing Date: September 30, 1983

Applicant: MITSUBISHI DENKI K.K.

2. Claims

(1) A speech recognition apparatus which extracts feature parameters of an input audio signal, and calculates a similarity between the extracted feature parameters and feature parameters of speech segments registered in advance, thereby to recognize a speech, the apparatus comprising

a beginning and terminating end detection means for detecting beginning and terminating ends of the input audio signal in order to define an audio signal region targeted for recognition,

wherein the beginning and terminating end detection means includes:

a CORMAX calculation means for calculating a CORMAX, which is a maximum value of an autocorrelation function of the input audio signal in a predetermined range; and

a beginning and terminating end determination means for determining beginning and terminating ends of the audio signal based on the CORMAX.

(2) The speech recognition apparatus according to Claim 1, wherein the beginning and terminating end

determination means determines a beginning end of the audio signal region on condition that the CORMAX exceeds a preset threshold thereof.

(3) The speech recognition apparatus according to Claim 1 or 2, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform with the CORMAX below a preset threshold thereof has lasted for a preset duration.

(4) The speech recognition apparatus according to Claim 1, wherein:

the beginning and terminating end detection means further includes a level calculating means for calculating a quantity corresponding to an amplitude of the input audio signal waveform, and

the beginning and terminating end determination means determines the beginning and terminating ends of the audio signal based on the CORMAX and an output of the level calculating means.

(5) The speech recognition apparatus according to Claim 1, wherein the level calculating means includes a means for calculating a power of the input audio signal.

(6) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a beginning end of the audio signal region on condition that a power resulting from calculation by the power calculating means exceeds a preset threshold thereof, and the CORMAX exceeds a preset



threshold thereof.

(7) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a beginning end of the audio signal region on condition that a power resulting from calculation by the power calculating means exceeds a preset threshold thereof, and a value resulting from multiplication of the CORMAX and a preset constant is larger than the power.

(8) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform has lasted for a preset duration when the input waveform meets at least one of a requirement that the CORMAX is below a preset threshold thereof and a requirement that a power resulting from calculation by the power calculating means is below a preset threshold thereof.

(9) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform has lasted for a preset duration when the input waveform meets at least one of a requirement that a power resulting from calculation by the power calculating means is below a preset threshold thereof and a requirement that a value resulting from multiplication of the CORMAX and a preset number is smaller than the power.